Forecasting triads: the negative feedback problem

Problem presented by

Stephen Custance-Baker

British Energy

Problem statement

A "Triad" is one of the three half-hour periods in the winter that have the highest national electricity demand. The charges levied on commercial users of electricity, for the *whole* of the winter, depend strongly on their consumption during Triad periods. In order to help its customers reduce these costs, British Energy (in common with other large electricity suppliers) attempts to forecast when Triads will occur and warn customers, who can then reduce their consumption during these periods. However, the very act of warning customers reduces the overall consumption and this may prevent a putative Triad from actually occurring. The Study Group was asked to consider whether this "negative feedback" problem could be prevented.

Study Group contributors

Joyce Aitchison (JMA Numerics)
John Billingham (University of Nottingham)
John Byatt-Smith (University of Edinburgh)
Liam Clarke (London School of Economics)
Jeff Dewynne (University of Oxford)
Kim Evans (University of Nottingham)
Vera Hazelwood (Smith Institute)
Sam Howison (University of Oxford)
Robert Hunt (University of Cambridge)

Report prepared by

Robert Hunt (University of Cambridge)
Liam Clarke (London School of Economics)
John Byatt-Smith (University of Edinburgh)
John Billingham (University of Nottingham)

1 Introduction

At the end of the winter season (November 1st to February 28th), National Grid identifies the three half-hour periods of highest consumption of electricity, subject to a separation of at least 10 calendar days. These three periods are called Triads. Thus, the half-hour in which national demand was highest is the first Triad; the second highest half-hour, excluding a period of 21 days centred on the first Triad, is the second Triad; and the third highest half-hour, excluding two periods of 21 days centred on the first and second Triads, is the third Triad. Historically, Triads always occur on Mondays to Thursdays between 16:30 and 18:00 hours, and not close to Christmas (which is a period of relatively low consumption). Figure 1 shows the typical pattern over a winter.

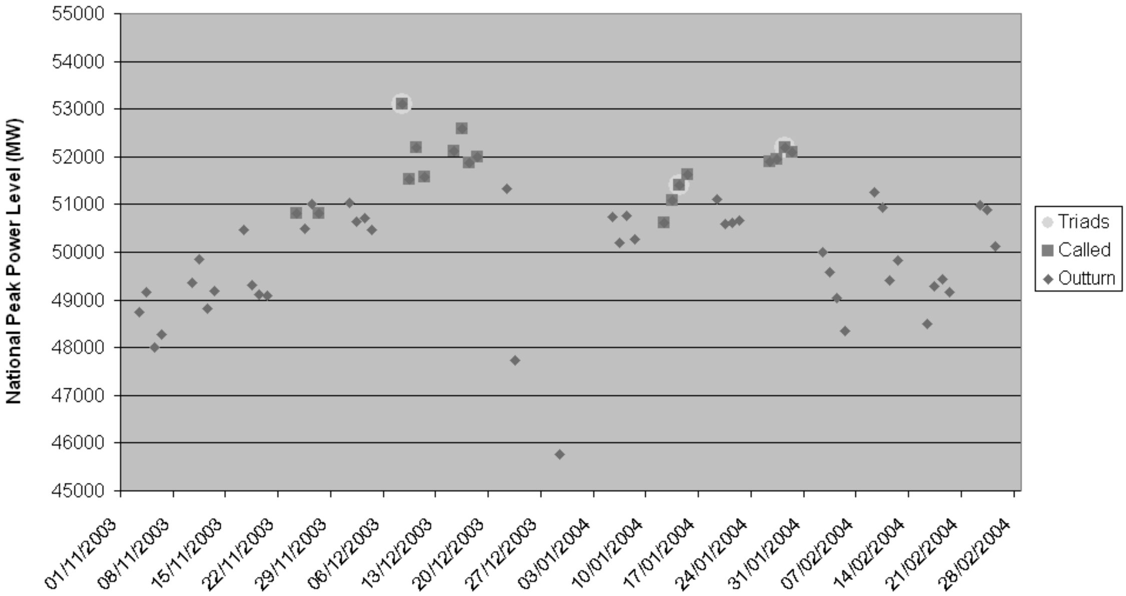

Figure 1: National electricity consumption figures from the winter of 03/04. Diamonds indicate daily levels; squares indicate dates on which British Energy issued Triad warnings; and circles indicate those dates on which a Triad was, at the end of the winter, declared to have occurred.

British Energy supplies electricity to large industrial and commercial customers. Charges are based on the Transmission Network Use of System (TNUoS) Charges which are levied by National Grid and passed through to British Energy's customers. TNUoS includes a substantial surcharge based on each customer's usage during the Triad periods: the higher a customer's usage during Triads, the higher their overall electricity supply costs.

British Energy, in common with other large commercial electricity suppliers, aims to reduce National Grid charges for some of its customers by issuing a "Triad warning" on days when it seems possible that a Triad might later be found to have occurred. Industrial customers are often encouraged by their contract with British Energy to reduce their consumption on those warning days. British Energy are restricted by the contract to issuing a limited number of calls over the entire winter period, anything up to 23.

A customer who receives a Triad warning may, or may not, take action (*e.g.* shutting down their factory early that day). Some – but *not* all – customers are contractually required to inform British Energy whether or not they are reducing their consumption. Therefore British Energy cannot be sure by how much overall consumption will drop if a warning is issued. Furthermore, the national consumption is also affected by suppliers other than British Energy who may also have issued a Triad warning.

It is thought that this system may suffer from "negative feedback". That is, on a day that is likely to include a Triad period because it has high predicted consumption, many suppliers will issue warnings, resulting in many customers reducing their actual consumption, ensuring that the total national demand is actually much lower than predicted: sufficiently low, in fact, that no Triad occurs, and no warning was actually necessary. It is expensive for customers to take action when it is not actually required.

British Energy currently uses the deterministic half-hourly consumption forecast issued by National Grid and the recent and forecast temperatures at hourly intervals at 7 locations around the country to decide on a daily basis whether to issue a Triad warning. The decision-making tool, called TriFoS, does not currently take into account the possible problem of negative feedback.

The Study Group was asked to consider ways of compensating for negative feedback. The next section confirms that feedback is a statistically significant phenomenon. Sections 3 and 4 then use the so-called 'Full-Information Secretary Problem' to motivate the derivation of possible triad-calling strategies. Section 5 examines the possibilities for issuing triad warnings from analysing historical data.

2 Detecting feedback

It is understood that a Triad warning issued by British Energy will reduce consumption by around 200 MW, whilst a warning issued by all suppliers (including British Energy) will reduce consumption on the order of 1200 MW.

An important step was to try to quantify the impact of negative feedback in Triad forecasting. We expect that the issuing of high-consumption warnings affects the behaviour of British Energy's commercial customers, encouraging them to reduce their consumption. In addition, we expect other supply companies to issue their own independent warnings or to react to British Energy's. To investigate this phenomenon we look at the forecast errors in National Grid's consumption forecast on those days when British Energy issued Triad warnings, and compare them to days when warnings were not issued. It is assumed that National Grid's own consumption forecasts do not take into account the effects of Triad warnings, in which case, we expect the forecast error to be larger on those days when warnings are issued.

We divide the days into 4 categories (which are not mutually exclusive), as follows.

- No Warning: days where no Triad warning was made by TriFoS;

- Low: days where Triad warnings were made and were deemed low probability by TriFoS;
- High: days where Triad warnings were made and were deemed high probability by TriFoS;
- All: days where Triad warnings were made regardless of probability.

Table 1 shows the mean and standard deviation of the consumption forecast error for these 4 categories. We see that in each case shown here, the mean forecast errors for days when warnings were issued (All) are larger than on those days when No Warning was issued. In addition, the mean forecast error for warning days with Low probability is, in general, larger than for those warning days with High probability. Whilst these observations are encouraging regarding the detection of feedback a more rigorous test is required in order to confirm our suspicions.

		No Warning			Low			High			All		
Period	N	n	μ	σ	n	μ	σ	n	μ	σ	n	μ	σ
01–02	109	90	702	661	12	1084	801	7	989	478	19	1049	685
02–03	101	78	759	1162	10	1152	563	13	753	454	23	926	532
03–04	94	76	562	503	10	1250	582	8	1091	378	18	1179	495
04–05	106	88	471	522	15	875	665	3	778	584	18	858	637
05–06	85	62	914	711	15	1055	449	8	1391	839	23	1172	615
01–06	495	394	668	757	62	1064	613	39	997	578	101	1038	598

Table 1: Summary of consumption forecast errors showing the period 2001–2006 (in years) and the number of data points. The data is subdivided into No Warning days, Low Triad-probability days, High Triad-probability days, and All Triad days (combining low and high probability days). The total number of points N, the number of data points of each type n, the mean μ and standard deviation σ of the forecast errors are presented.

Table 2 shows the P-values from a Student t-test on the difference of the means between No Warning and each of Low, High and All. The null hypothesis is that there is no difference in mean forecast error. This is tested against the alternative hypothesis that the mean error is larger for Low, High and All than for No Warning. The test is done assuming unequal variances in the samples. The values given are the probabilities of observing the given result, or one more extreme, by chance if the null hypothesis is true. Small values of P cast doubt on the validity of the null hypothesis.

The test shows clearly that, taking the data over all years, we can reject the null hypothesis that the mean forecast errors are the same. Therefore, there is clear evidence that Triad warnings *do* affect the measured consumption, and negative feedback is a real effect.

The TriFoS system used by British Energy performs very well, capturing almost all Triads in the given number of warnings. Indeed, the current maximum number of warnings (23) seems a little generous: in all the years for which TriFoS has been in operation, a considerably smaller number would have been sufficient.

	P-value		
Period	Low	High	All
01–02	0.069	0.089	0.027
02–03	0.046	0.514	0.167
03–04	0.002	0.002	0.000
04–05	0.020	0.230	0.012
05–06	0.171	0.080	0.053
01–06	0.000	0.001	0.000

Table 2: P-values from one-sided t-test on consumption forecast error. The values provided are the probabilities of the corresponding t-statistic.

3 Approaches based on the Secretary Problem

In the Full-Information Secretary Problem (or FISP), one imagines being presented with a sequence of n values, which are drawn independently from some known probability distribution.[1] One selects (or 'calls') r of the n values, in an attempt to include among the selection the M largest values (assuming always that $r \geq M$). However, each value may be selected only when it is presented, *i.e.* without knowledge of the values yet to come, and without going back to any values that have been previously passed over. The objective is to devise a strategy that maximises the probability of including the M largest values among the r selections. The problem is known for short as the '(r, M) FISP'.

The similarity with the calling of Triads should be clear, with the data values being estimates of daily peak power consumption. However, the Triad-calling problem has two important additional features, namely that energy consumption data is:

(1) *Correlated.* The best way of predicting the weather today is to guess that it will be the same as yesterday. Similarly, electricity consumption on a given day is not independent of energy consumption on the previous day.

(2) *Subject to negative feedback.* If British Energy call a Triad, many of their customers will reduce their consumption, as will commercial customers of other electricity suppliers who provide the same service, thereby reducing the consumption below that predicted.

In order to investigate the effects of these two complications, we will study the FISP with correlation and feedback. We will not, in this report, attempt to include the effect of the ten-day window or to add any uncertainty into the prediction, but these points are discussed further in section 6.

Appendix A reviews the optimal strategy for the $(1, 1)$ FISP, as discussed by Gilbert and Mosteller [2], which takes the following form. Suppose that the n data values

[1] The term 'full information' refers to the fact that the underlying distribution of data values is known.

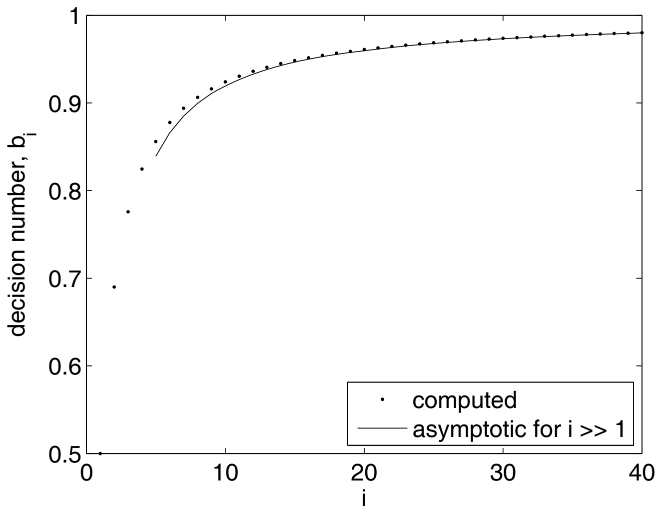

Figure 2: The decision numbers for the Full-Information Secretary Problem with draws from a uniform distribution, and their asymptotic approximation for $i \gg 1$.

are $y_1, y_2, \ldots, y_n$. Then the best strategy involves an increasing sequence of **decision numbers** $\{b_i\}$, which does not depend on n. The ith data point should be called if

$$y_i > b_{n+1-i} \qquad \text{and} \qquad y_i = \max(y_j, j \leq i)\,. \tag{1}$$

In other words, a value is called if it is both the largest value seen so far and larger than the decision number b_j, where j is the number of values yet to come (including the current value). Note that $b_1 = 0$, since we always call the final data point if it the largest in the whole set.

The FISP is most easily studied if the data points are drawn from the uniform distribution on [0 1], $U(0,1)$, and for the $(1,1)$ FISP we then have (see the appendix)

$$b_{k+1} = \frac{1}{1+z}\,, \text{ where } 1 = \sum_{j=1}^{k} \frac{1}{j} \begin{pmatrix} k \\ j \end{pmatrix} z^j. \tag{2}$$

Note that $b_2 = 0.5$. It is straightforward to compute the higher decision numbers numerically and asymptotically they increase to 1 according to

$$b_{i+1} \sim 1 - \frac{0.80435}{i} \quad \text{as } i \to \infty. \tag{3}$$

The decision numbers, along with this asymptotic result, are shown in figure 2. The decision numbers b_i are monotonically increasing with i. In practice, the optimal strategy for n large is to wait, unless a data point very close to unity appears, but, as the end of the data set approaches and the decision numbers get smaller, the next data point to come along that is the largest so far should be called. As discussed in [2], the probability of calling the maximum correctly using this strategy, $P(\text{win})$, is about 0.61 when $n = 10$, and 0.58 when n is large.

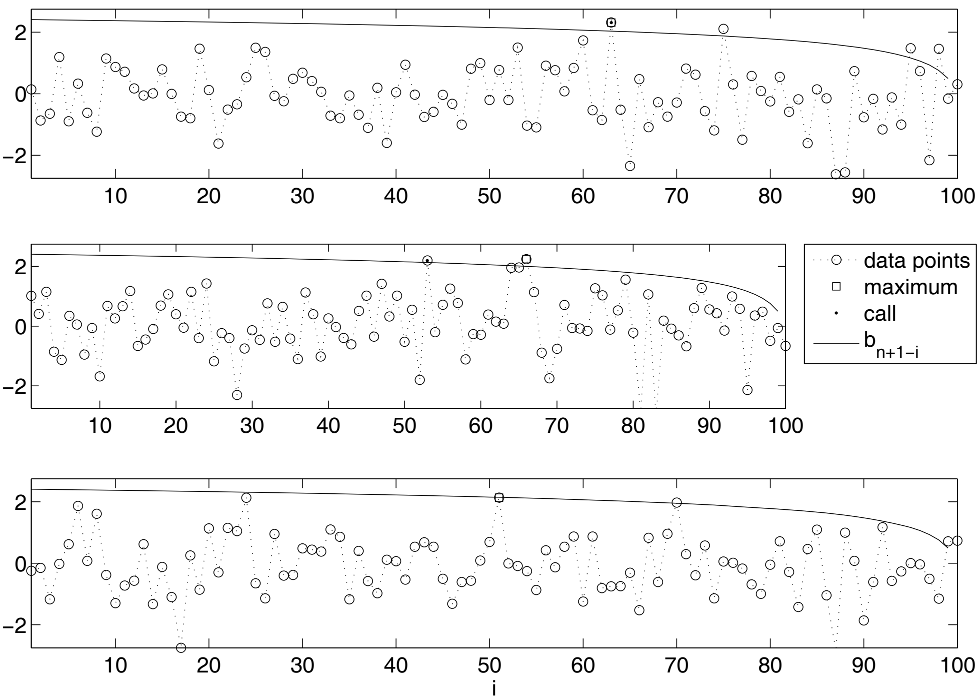

Figure 3: Examples of the three possibilities in the $(1, 1)$ FISP with $n = 100$: the maximum is called correctly (top figure), the wrong data point is called (middle figure), or the final point is reached without a call being made (bottom figure).

Other underlying probability distributions for the data values can be accommodated through mapping to $U(0, 1)$ through the cumulative distribution function, $F(y)$. Provided that $F(y)$ is continuous and strictly increasing then setting $x_i = F(y_i)$ generates values x_i that are drawn from $U(0, 1)$. For example, a draw y from the normal distribution with mean zero and unit variance, $N(0, 1)$, can be mapped to a draw x from $U(0, 1)$ using

$$x = \frac{1}{2}\left\{1 + \operatorname{erf}\left(\frac{y}{\sqrt{2}}\right)\right\} \tag{4}$$

where erf is the ”error function” obtained through integrating the normal distribution. We would expect a reasonable model for electricity consumption data to involve normally distributed random variables, which can then be mapped to a uniform distribution in this manner.

The decision numbers on $U(0, 1)$ must be mapped to those for $N(0, 1)$ using (4) to obtain the correct numbers for the normally distributed version of the problem. Figure 3 shows the decision numbers and some typical sequences of data points for normally distributed FISP. Note that the horizontal axis is inverted compared to Figure 2, so that it indicates the number of points that have been seen, rather that number still to be seen.

3.1 Framework for incorporating correlation and feedback

In order to model electricity consumption data, we wish to extend the FISP, with normally distributed data, to take account of correlation and negative feedback. The model we use assumes a sequence of n data points of the form

$$y_{i+1} = (1-\alpha)\, y_i + N(0,1), \quad i = 1,\, 2,\, \ldots,\, n-1, \tag{5}$$

with $y_1 = N(0,1)$. This means that the series is correlated for $O(1/\alpha)$ data points. With $\alpha = 1$, this is just the previous case of an uncorrelated draw from $N(0,1)$, and with $\alpha = 0$ it is a random walk. We will study the correlated, (r, M) FISP based on (5). The Triad prediction problem is closely related to this problem with $M = 3$ (and $r = 23$). In addition, we will assume that, when a call is made, negative feedback reduces y_i to $y_i - f$, with $f \geq 0$. We seek the optimal strategy as a function of α, f, r and M.

The random walk without feedback, corresponding to $\alpha = 0$ and $f = 0$, was studied in [3], and we discuss this case in section 3.2. We then extend the analysis to find analytically the optimal strategy for the $(1,1)$ uncorrelated FISP with feedback, and an almost optimal strategy for the $(r,1)$ uncorrelated FISP with no feedback. For the general problem, we determine strategies using Monte Carlo simulation, in Section 4.

3.2 Optimal strategy for the $(1,1)$ FISP with a random walk

The secretary problem for a random walk, (corresponding to $(r, M, \alpha, f) = (1, 1, 0, 0)$ in the general framework) was studied in [3], where it was shown that the optimal strategy is to observe the first k data points, and then pick the next data point that is larger than all the previous points. This is similar to the standard, zero information secretary problem, see [2], where the optimal strategy is the same, and $k \sim n/e$ as $n \to \infty$. However, for the random walk, this strategy is optimal for all k. The proof involves rearrangements of the sequence of data points, and is not given here. In particular, choosing the first data point, and choosing the final data point are both optimal strategies. It is also shown in [3] that the probability of winning using any of these strategies is asymptotic to $1/\sqrt{\pi n}$ for n large. Finally, we note that, as long as[2] $b_1 = -\infty$, so that the final data point is chosen if it is a maximum, any sequence of decision numbers b_i, possibly including $\pm\infty$, gives an optimal strategy.

3.3 Optimal strategy for the $(1,1)$ FISP with feedback

This case has $\alpha = 1$ and $f > 0$. We can adapt the method used by Gilbert and Mosteller, [2], described in appendix A, to incorporate feedback. Consider the position when the current data point, chosen from $N(0,1)$, is y, we have yet to call a maximum, there are k data points still to come, and $y - f$ is the largest value to appear so far. If we

[2] Recall that zero in $U(0,1)$ maps to $-\infty$ in $N(0,1)$.

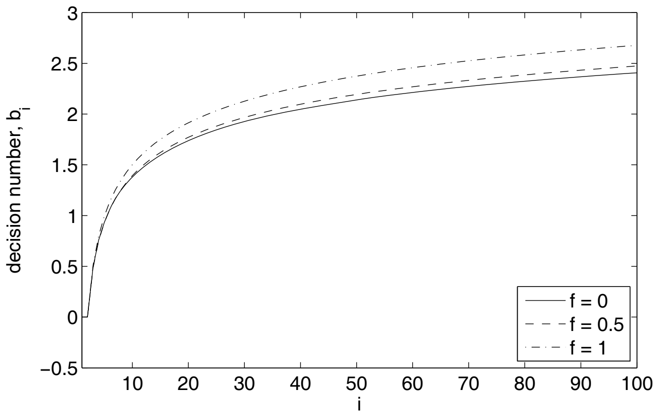

Figure 4: The decision numbers for the Full-Information Secretary Problem with data taken from $N(0,1)$, with feedback.

call a maximum, the probability that we are correct is equal to the probability that the remaining k data points are less than $y-f$, which is $(\Phi(y-f))^k$, where

$$\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2}s^2} ds \equiv \frac{1}{2}\left\{1 + \mathrm{erf}\left(\frac{y}{\sqrt{2}}\right)\right\}$$

is the cumulative density function for $N(0,1)$.

If we don't call a maximum and continue, and j of the remaining data points are greater than y, we will correctly call a subsequent maximum if the largest remaining data point is $y_m > y+f$ and the remaining $j-1$ points are less than $y_m - f$. The probability of calling the maximum correctly in this case can therefore be written as

$$\int_{y+f}^{\infty} \left\{\Phi(y_m - f) - \Phi(y)\right\}^{j-1} P(y_m)\, dy_m,$$

where

$$P(y_m) \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_m^2},$$

is the point probability of drawing y_m from $N(0,1)$. The appropriate decision number for the current data point is the value of y at which the probability of winning by calling a maximum and the probability of winning by not calling a maximum are equal. This gives us

$$(\Phi(y-f))^k = \sum_{j=1}^{k} \begin{pmatrix} k \\ j \end{pmatrix} (\Phi(y))^{k-j} \int_{y+f}^{\infty} \left\{\Phi(y_m - f) - \Phi(y)\right\}^{j-1} P(y_m)\, dy_m. \qquad (6)$$

Note that when $f = 0$, (6) is equivalent to (2) with $z = (1-\Phi(y))/\Phi(y)$, since $P = d\Phi/dy$.

It is straightforward to solve (6) numerically, and figure 4 shows the decision numbers for various values of f. We can see that, as we might expect, the decision numbers increase with f.

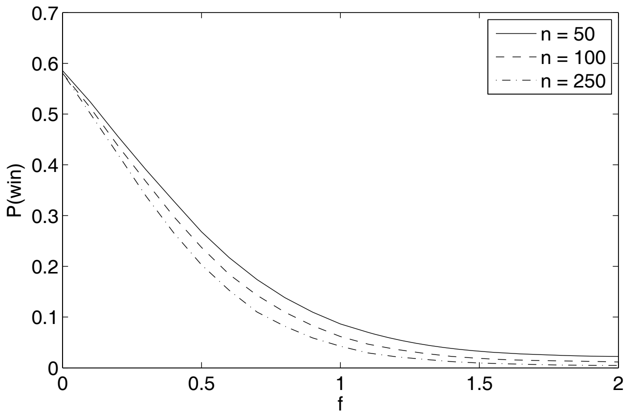

Figure 5: The probability of winning in the Full-Information Secretary Problem with feedback f, for various n.

We have not attempted to calculate the probability of winning analytically, but have estimated it using Monte Carlo simulation with 10^5 trials. Figure 5 shows the probability of winning as a function of f. Clearly, feedback has a significant effect on the probability of calling the maximum, which decreases monotonically with f. In addition, although $P(\text{win})$ asymptotes to a constant value as $n \to \infty$ when $f = 0$, this does not appear to be the case for $f > 0$, and we conjecture that $P(\text{win}) \to 0$ as $n \to \infty$.

3.4 An almost-optimal strategy for the $(r, 1)$ FISP

We return now to the case with no correlation and no feedback, but allow several calls to be made. We can extend the method used by Gilbert and Mosteller, [2], described in appendix A, which works when $r = 1$, to find a strategy for uncorrelated FISP with $r > 1$. Once we have a strategy for the cases $r < S$, then we have a strategy for the case $r = S$ once a single call has been made. In particular, we look for a strategy involving decision numbers b_i^R, for $i = 1, 2, \ldots, n$ and $R = 1, 2, \ldots, r$, such that a call should be made if, with R calls remaining in hand, $y_i > b_{n+1-i}^R$ and $y_i = \max(y_j, j \leq i)$. It is easiest here to work with data points taken from the uniform distribution, $U(0, 1)$, transforming the results as required to other distributions.

Consider the position when the current data point, x, is the largest so far, we have R calls available, and there are k data points still to come. If we call a maximum, the probability that either x is the maximum or we will still be able to call the maximum with one of our $R - 1$ remaining calls, is composed of two parts, say $P_1 + P_2$. These are, firstly, the probability that no more than $R - 1$ of the remaining k data points are greater than x,

$$P_1 = \sum_{j=0}^{R-1} \binom{k}{j} x^{k-j} (1 - x)^j ,$$

and secondly, the probability P_2 that, if there are $j \geq R$ data points remaining that are bigger than x, then they are ordered so that the $R - 1$ remaining calls suffice to find the maximum. For example, consider the situation when $R = 5$ and $j = 6$, and the 6 remaining data points bigger than x are $x_6 > x_5 > x_4 > x_3 > x_2 > x_1 > x$. Then, if these data points are ordered, for example, $x_1x_2x_3x_4x_5x_6$, the $R - 1 = 4$ remaining calls will be used up on the first 4 data points, and x_6, the maximum, will not be called. However, if they are ordered, for example $x_2x_3x_1x_4x_6x_5$, the data points x_2, x_3, x_4 and x_6 will be called, which includes the maximum. The data point x_1 is effectively shielded by the data point $x_2 > x_1$, which appears earlier in the sequence.

In general, if the maximum data point is the Ith that is larger than x, it is the ordering of the $I-1$ preceding data points that concerns us. Let N_γ^β be the number of permutations of $x_1 < x_2 < \ldots < x_\beta$ such that our decision process would have to make a total of γ calls to reach the end of the sequence. (That is, γ is the smallest number of calls needed to include the maximum x_β as one of the calls.) For the two examples given above, $\beta = 6$. In the first case $\gamma = 6$ calls would be made, and in the second, $\gamma = 4$ calls would be made. We can find a recurrence relation for N_γ^β by noting that all sequences of length $\beta+1$ can be made by adding a new data point, $x_0 < x_1$, to each sequence of length β, in every possible position. The number of calls increases by one if the point is added to the left hand end of the sequence, but remains unchanged otherwise. For example, starting from the sequence $x_2x_3x_1x_4x_6x_5$, which leads to 4 calls, and adding $x_0 < x_1$, we obtain

$$\left.\begin{array}{l} x_2x_3x_1x_4x_6x_5x_0, \\ x_2x_3x_1x_4x_6x_0x_5, \\ x_2x_3x_1x_4x_0x_6x_5, \\ x_2x_3x_1x_0x_4x_6x_5, \\ x_2x_3x_0x_1x_4x_6x_5, \\ x_2x_0x_3x_1x_4x_6x_5, \end{array}\right\} \text{4 calls required,} \qquad x_0x_2x_3x_1x_4x_6x_5, \quad \text{5 calls required.}$$

This example shows how N_γ^β affects both the values $N_\gamma^{\beta+1}$ and $N_{\gamma+1}^{\beta+1}$. This means that

$$N_\gamma^{\beta+1} = N_{\gamma-1}^\beta + \beta N_\gamma^\beta,$$

which, along with the condition $N_0^0 = 1$, $N_\gamma^0 = 0$ for $\gamma > 0$, completely defines N_γ^β. However, this is precisely the definition of the unsigned Stirling numbers of the first kind (see [1] for a discussion), and the usual notation is

$$N_\gamma^\beta \equiv \left[\begin{array}{c} \beta \\ \gamma \end{array} \right].$$

The mathematical details of this approach (P_2 and an equation for the decision numbers) are given in appendix B.

It is straightforward to solve for the decision numbers numerically, and figure 6 shows them up to $r = 10$. As we might expect, the decision numbers become smaller as r increases, since, with more calls available, we can use them more liberally. It is also easy to show that $b_{i+1} = 1 + O(i^{-1})$ for $i \gg 1$, but we omit the details here. We can now see why this strategy is not optimal. We constructed it by assuming that we would always choose a data point greater than the largest called so far, since the decision numbers b_{n+1-i} decrease monotonically with i. However, after making a call with R calls in hand,

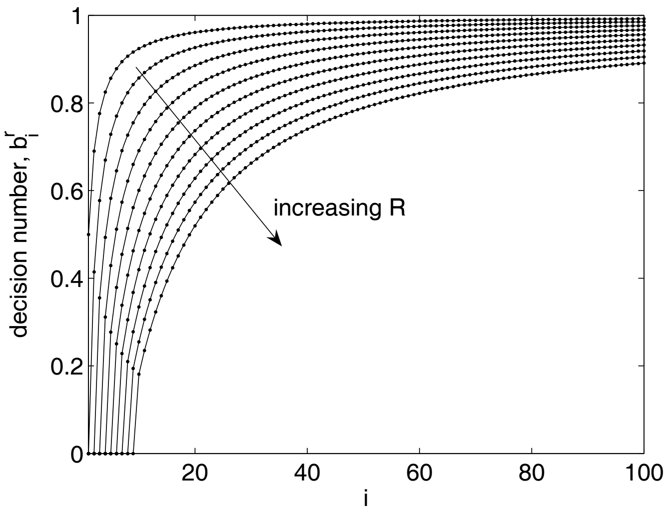

Figure 6: The decision numbers for the Full-Information Secretary Problem with draws from a uniform distribution and r calls, for r up to 10.

we switch to the decision numbers appropriate for $R - 1$ calls in hand, which means that the decision number increases discontinuously after each call. There is therefore the possibility of a data point arriving that is bigger than the largest point called so far, but smaller than the appropriate decision number, contradicting the assumptions that underlie the construction. We shall however see in the next section that our strategy is close to optimal.

Finally, figure 7 shows the probability of calling the maximum as a function of r when $n = 100$. The probability of calling the maximum increases with r, and is greater than 0.99 when $r = 5$. Also shown is the probability of winning using the simpler strategy suggested by Gilbert and Mosteller, [2], which uses a sequence of decision numbers that depends upon r, but is independent the number of calls made. As we would expect, the strategy derived here gives a slightly higher probability of calling the maximum.

4 Optimal strategies by Monte Carlo simulation

In order to investigate the cases for which we cannot derive an optimal strategy analytically, we will use Monte Carlo simulation. Before outlining the procedure, we need to take note of two features of the problem.

Optimal strategy with feedback and more than one call ($f > 0$, $r > 1$)

With just one call available, $r = 1$, we showed in section 3.3 that we can derive the optimal strategy. In particular, it is clear that we should not make our call until the current data point exceeds by more than f the largest data point to appear so far, since feedback will otherwise reduce it too far for it to remain a maximum. When $r > 1$, the

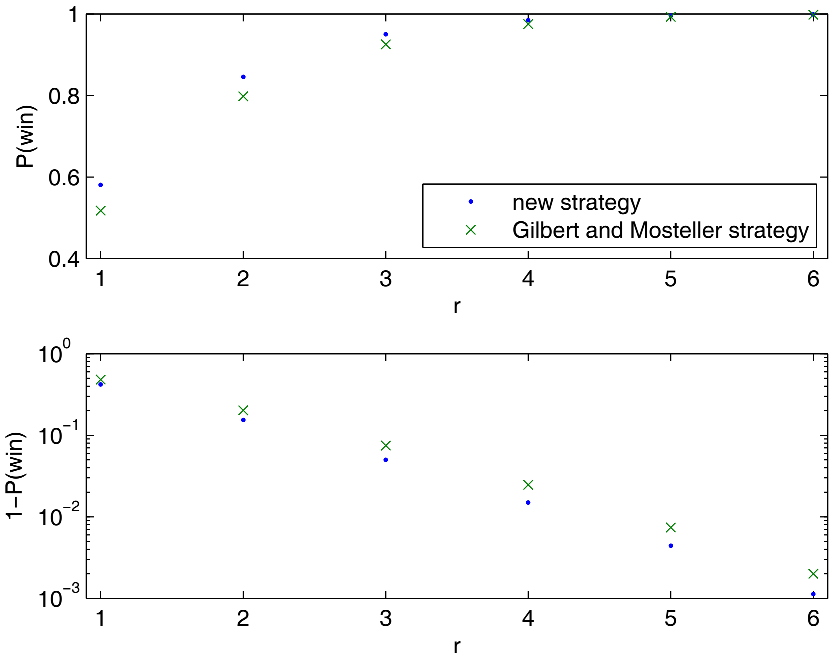

Figure 7: Probability of calling the maximum with r calls, from Monte Carlo simulation with 10^5 trials and $n = 100$. Also shown is the strategy due to Gilbert and Mosteller [2].

situation is not so clear. If we use this approach at every data point, we may discard a sequence of increasing data points, and push ever higher the threshold above which we will actually call. An alternative strategy is, if we have more than one call still available, to make a call whenever the current data point exceeds the greatest value not called. Although we know that we will not necessarily call a maximum, we will reduce the value of the called point by f, so that the threshold does not increase. We can then, with one of our remaining calls, hope to call a data point that is large enough that, even with the effect of feedback, it is actually a maximum. This is the strategy that we adopt below. However, we might envisage cases where the current data point only exceeds the previous maximum by a small amount, in which case we would not want to waste a call on it. This suggests that an optimal strategy involves an additional sequence of numbers, $\{\delta_i^R(f)\}$, such that with i data points and R possible calls remaining, we call if the current data point exceeds the previous maximum by more than δ_i^R. We might expect that $\delta_i^R \ll 1$ when $R \gg 1$ and that $\delta_i^1 = f$. We have yet to investigate this more sophisticated strategy further. In effect, we assume that $\delta_i^1 = f$ and $\delta_i^R = 0$ for $R > 1$.

Optimal strategy for calling more than one maximum ($M > 1$)

When the $M > 1$ largest data points are to be called, the optimal strategy is very difficult to calculate, since the decision number depends not only on how many data points remain, but also the relative sizes of the previous data points and, once a call has been made, the size of the data points called. As an example, we can study the

$(2, 2)$ FISP. Consider the situation when we have yet to make a call. If we try to use the method described earlier, we must calculate the probability of calling the two largest data points, firstly if we call the current data point, and secondly if we do not call the current data point. If we make a call, the probability of winning is the probability that, of the remaining k data points, just one is bigger than the second highest data point seen so far, which is not the current data point, but the previous highest data point. In other words, the optimal strategy must depend upon not only the current data point, but also the previous highest data point. The decision numbers are therefore functions not just of i and R, but also of, in this example, the previous highest data point, and in general, the previous $M - 1$ highest data points. It may be possible to sort this out in principle, and even in practice, but we have not attempted to do so here. Instead, we look for decision numbers $b_i^{(R,M)}$ that are dependent only upon i, R and M, as described below.

4.1 Monte Carlo simulations with $M = 1$

4.1.1 Numerical method

In principle, when successive data points are correlated, we could use the methods described in section 3 to calculate the decision numbers. However, in practice, it is not possible to determine analytically the probabilities of winning by calling or not calling a given data point, because of the complicated conditional probabilities that need to be calculated. Instead, we calculate the decision numbers numerically using Monte Carlo simulation. Specifically, if we know the decision numbers b_i^r for $1 \leq i < j$ and b_i^R for $1 \leq i \leq n$ and $1 \leq R < r$, we can calculate b_j^r numerically. Using $n = j$, we can calculate the probabilities of winning by calling the first data point, P_{c}, and by not calling it, P_{nc}, using Monte Carlo simulation, evolving y_i using (5), including the effect of feedback if necessary, since we know the decision numbers for the subsequent $j - 1$ data points. In this manner we can compute $G(y_1) \equiv P_{\mathrm{c}} - P_{\mathrm{nc}}$. By construction, b_j is the solution of $G(y_1) = 0$. Since we compute G by Monte Carlo simulation, we cannot rely on it being smooth, so we use a bisection method to find the root. This works well in practice until the probabilities P_{c} and P_{nc} both become either small or close to unity. In these cases, the function G eventually becomes too noisy to work with. In the results that follow, we have used Monte Carlo simulation with 10^5 trials to estimate G. Finally, note that, since we know that $b_i^R = -\infty$ for $1 \leq i \leq R$ (the final R data points should be called if they are maxima), we have a place to start our calculation, and therefore a practical algorithm for calculating the decision numbers.

4.1.2 Validation for the $(r, 1)$ FISP and $(1, 1)$ FISP with feedback

We begin by validating our numerical method against the analytical results that we obtained earlier. Figure 8 shows the decision numbers for the $(r, 1)$ FISP calculated both numerically, by Monte Carlo simulation, and analytically using the results of section 3.4. Although, as we discussed above, the analytical decision numbers, shown

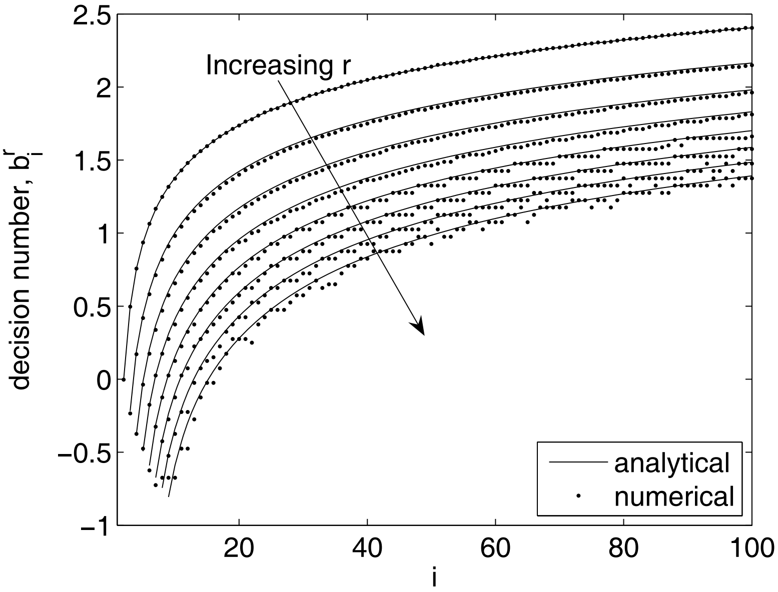

Figure 8: The decision numbers for the $(r, 1)$ Full-Information Secretary Problem, as given by the solutions of (11) (solid lines), and by Monte Carlo simulation (points).

as solid lines, cannot be precisely optimal, they are in good agreement with the optimal decision numbers calculated by Monte Carlo simulation, which are slightly lower. In particular, the strategy when $r = 1$ (the standard $(1, 1)$ FISP) is in excellent agreement. We conclude that the analytically determined strategy is close to optimal. Note that, as r increases past five, the probability of winning gets close enough to unity that our method for determining the decision numbers numerically loses accuracy, which leads to the noise in the calculated decision numbers evident in figure 8.

Figure 9 shows a comparison between the analytically and numerically calculated decision numbers for $(1, 1)$ FISP with feedback. Again there is good agreement.

4.1.3 Results for the $(r, 1)$ FISP with correlation and feedback

Now that we have some confidence in our Monte Carlo simulations, we can investigate the effect of correlation and feedback. Figure 10 shows the decision numbers for $(1, 1)$ FISP with correlation. As α (in Equation 5) decreases, the decision numbers become larger, which reflects the tendency of the data points to wander further from zero. When $\alpha = 0.01$, the correlation extends over about 100 data points, and the series is close to a random walk. This is why the decision numbers are so noisy in this case, since, as we saw in section 3.2, when $\alpha = 0$, any sequence of decision numbers with $b_1 = -\infty$ provides an optimal strategy. This is consistent with figure 11, which shows the probability of calling the maximum as a function of α. The circle shows the analytical result, $P(\text{win}) \sim 1/\sqrt{\pi n}$ for $n \gg 1$ and $\alpha = 0$. Figure 12 shows the decision numbers for $\alpha = 0.5$ and various r. As r increases, the decision numbers become more noisy, again reflecting the fact that the probability of calling the maximum becomes close to unity. The probability of

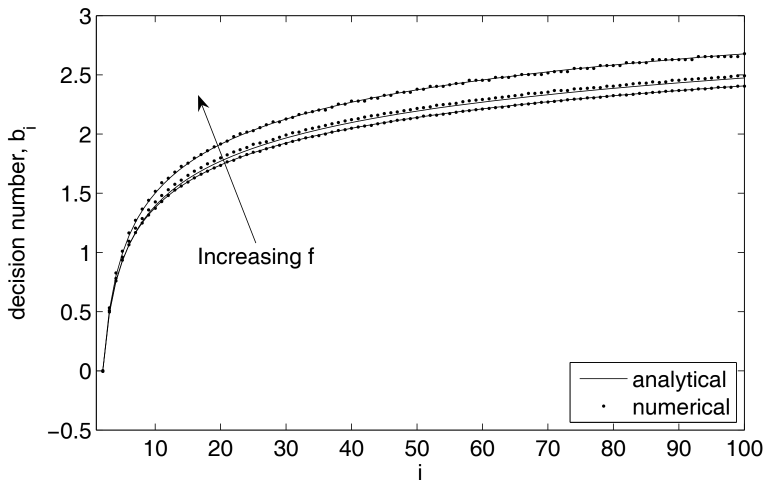

Figure 9: The decision numbers for the $(1, 1)$ Full-Information Secretary Problem with feedback, as given by the solutions of (6) (solid lines), and by Monte Carlo simulation (points).

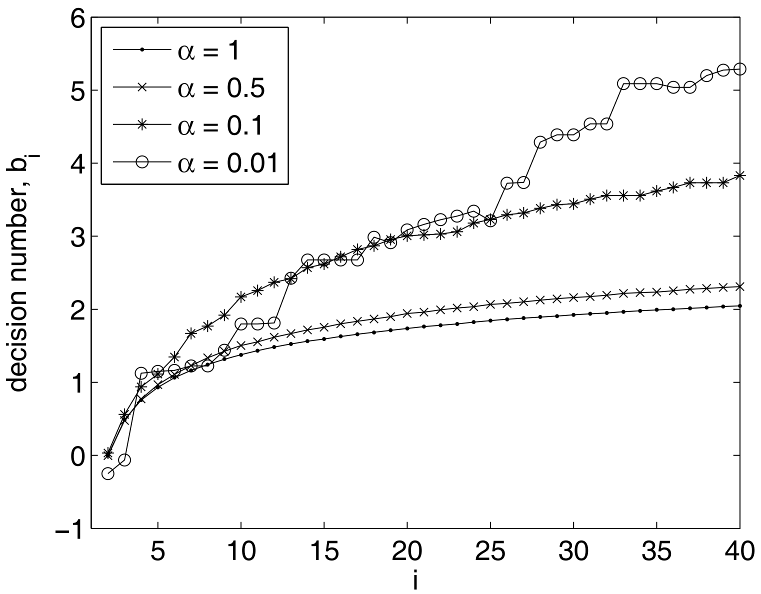

Figure 10: The decision numbers for the $(1, 1)$ Full-Information Secretary Problem with correlation, but no feedback ($f = 0$), calculated by Monte Carlo simulation.

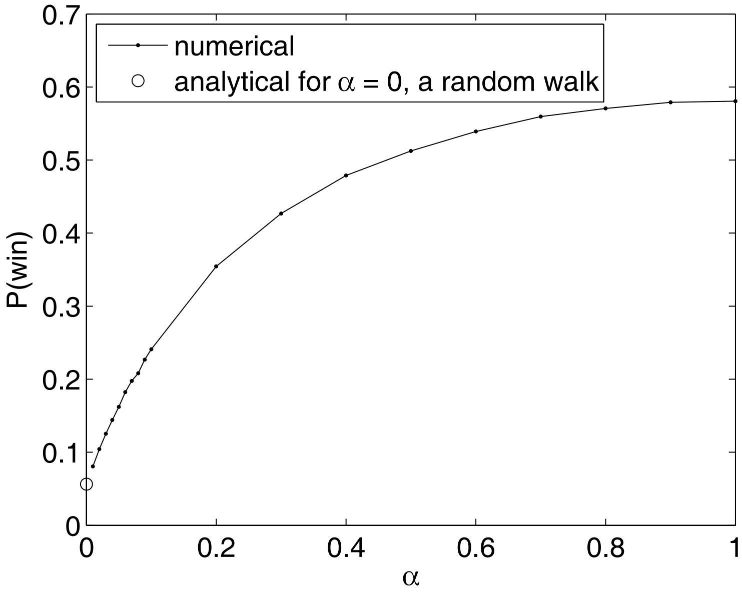

Figure 11: The probability of calling the maximum for the $(1, 1)$ Full-Information Secretary Problem with correlation, but no feedback ($f = 0$), and $n = 100$ data points, calculated by Monte Carlo simulation.

calling the maximum is shown in figure 13 for various values of r as a function of α. It is clear that, even with correlation, the probability of winning rapidly approaches unity as r increases, unless α is close to $1/n$, which is equal to 0.01 in this case.

Finally, we present some results with $f = 0.5$. Figure 14 shows the decision numbers when $f = 0.5$ and the data points are uncorrelated ($\alpha = 1$). One striking feature of this plot is the gap between the decision numbers for $r = 1$ and those for $r = 2$, which is significantly larger than the gaps between the remaining sets of decision numbers. This is probably because the optimal strategy involves the additional sequence $\{\delta_i^r\}$, described earlier. In our simulations, we call a maximum whenever a new largest point arrives, even if we know that feedback will reduce it below the threshold, until we have just one call remaining, when we have to exercise more caution. The probability of calling the maximum using these decision numbers is shown in figure 15, and is significantly smaller for a given r than when there is no feedback (see figure 13). Figures 16 and 17 show the results when $f = 0.5$ and $\alpha = 0.2$. Although the decision numbers are somewhat larger than those shown in figure 14, the probabilities of calling the maximum shown in figure 17 are not very different from those when there is no correlation, shown in figure 15, which suggests that feedback is the dominant mechanism that determines the probability of success.

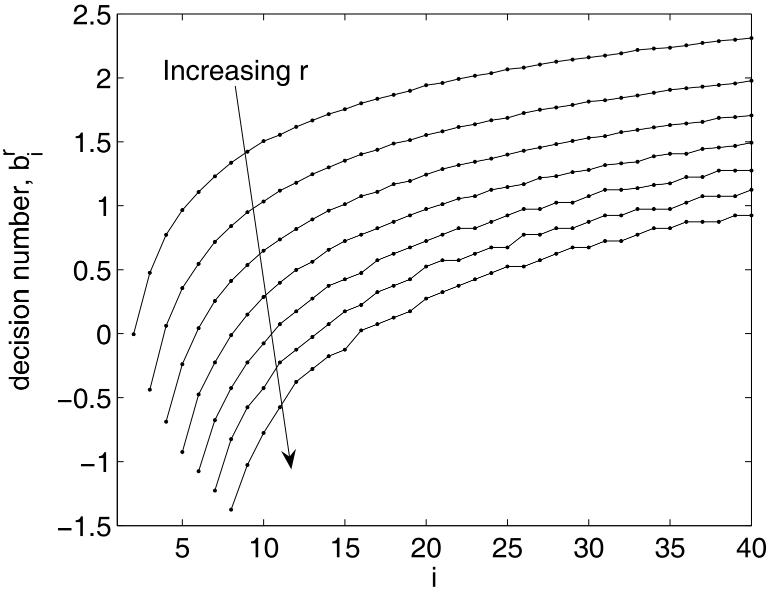

Figure 12: The decision numbers for the $(r, 1)$ Full-Information Secretary Problem with correlation $\alpha = 0.5$, but no feedback ($f = 0$), calculated by Monte Carlo simulation.

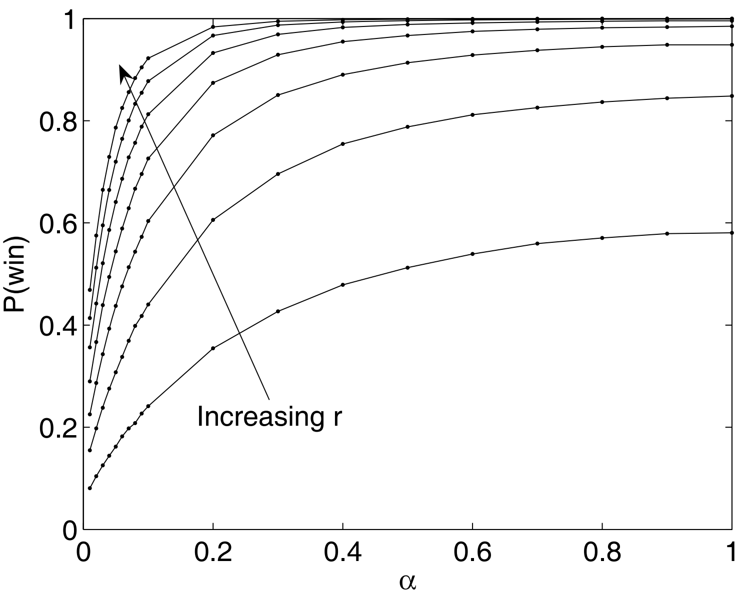

Figure 13: The probability of calling the maximum for the $(r, 1)$ Full-Information Secretary Problem with correlation, but no feedback ($f = 0$), and $n = 100$ data points, calculated by Monte Carlo simulation.

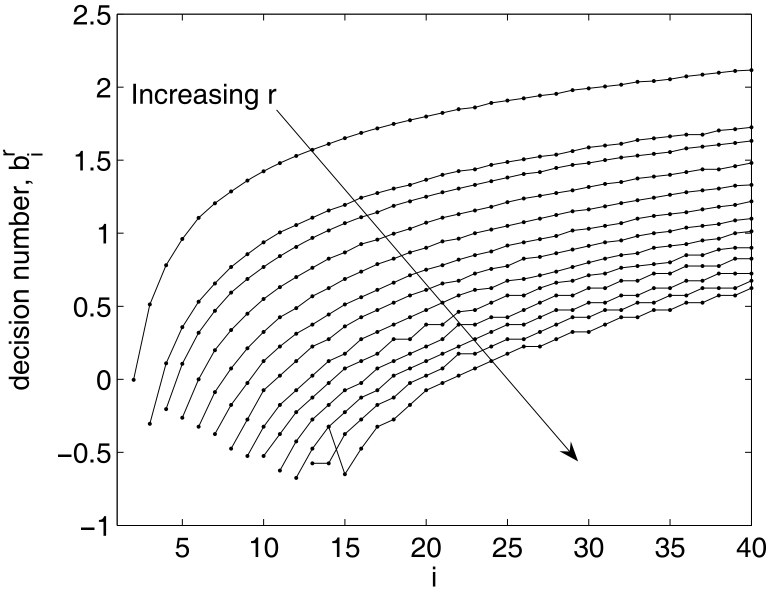

Figure 14: The decision numbers for the $(r, 1)$ Full-Information Secretary Problem with no correlation ($\alpha = 1$), and feedback $f = 0.5$, calculated by Monte Carlo simulation.

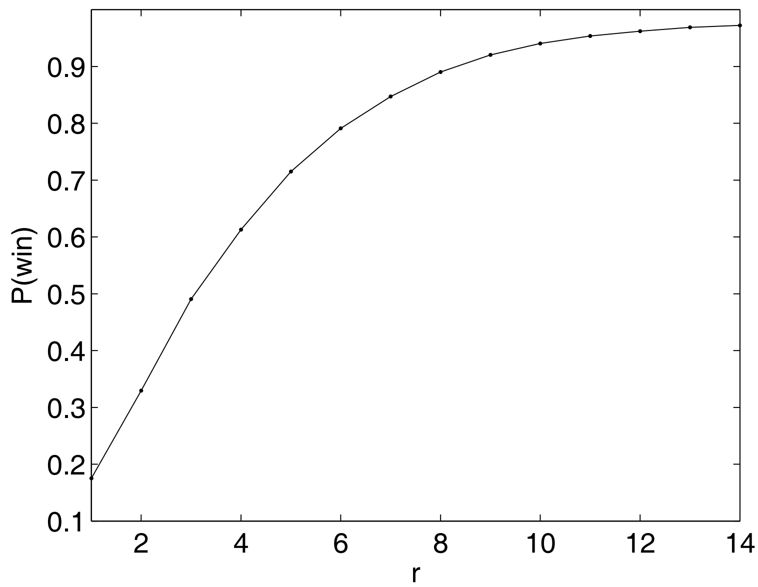

Figure 15: The probability of calling the maximum for the $(r, 1)$ Full-Information Secretary Problem no correlation ($\alpha = 1$), and feedback $f = 0.5$, and $n = 100$ data points, calculated by Monte Carlo simulation.

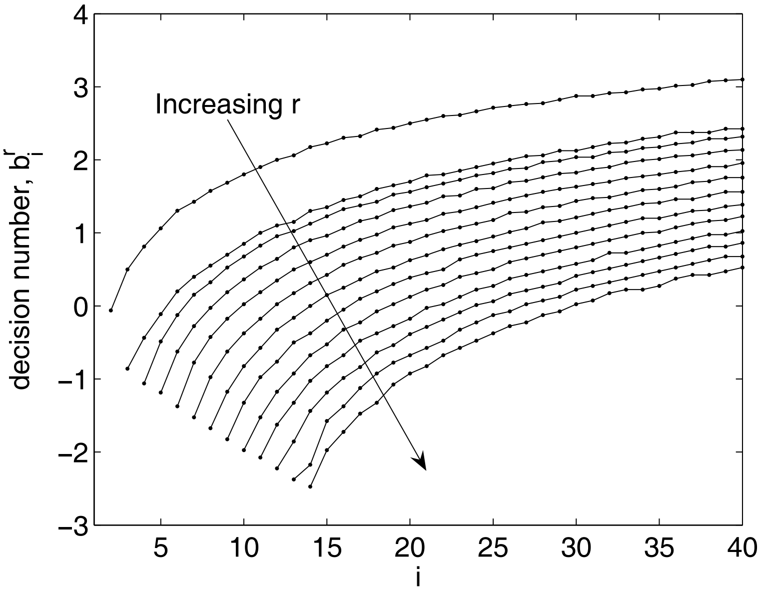

Figure 16: The decision numbers for the $(r, 1)$ full information secretary problem with correlation $\alpha = 0.2$, and feedback $f = 0.5$, calculated by Monte Carlo simulation.

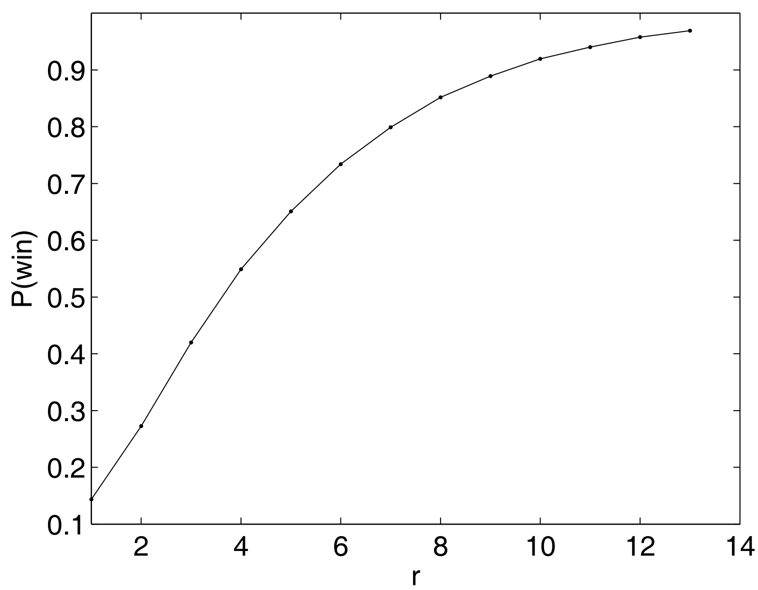

Figure 17: The probability of calling the maximum for the $(r, 1)$ full information secretary problem correlation $\alpha = 0.2$, and feedback $f = 0.5$, and $n = 100$ data points, calculated by Monte Carlo simulation.

4.2 Monte Carlo simulations with $M > 1$

As discussed earlier, the optimal strategy in this case depends explicitly on the previous data points. We will not attempt to unravel this here, but instead propose a simpler strategy. We will assume that there are sequences of decision numbers, $\{b_i^{(r,m)}\}$, with $1 \leq m \leq M$, which govern the selection of the first M calls when m calls remain to be made. However, once we have called at least M data points, we make another call only if the current data point is larger than the Mth largest of those already seen. It seems likely that this strategy is close to optimal since, as we saw above, our analytically determined strategy for the $(r, 1)$ FISP is close to optimal, even though it was constructed on precisely this basis (with at least one call made, call if the current point is the largest so far). We can calculate the decision numbers numerically in the same way as we did for the case $M = 1$. However, for these calculations, we used just 10^4 trials, since the algorithm required to sort out the position of the current data point relative to those called so far is rather time consuming, so that the accuracy of the following results is somewhat lower than those described above.

4.2.1 Results for the (r, M) FISP

Figure 18 shows the decision numbers when $r = M$, with neither correlation nor feedback. Note that, for example, the decision numbers for the $(3, 3)$ FISP are given by the third largest set in figure 18, but, once a call has been made, switches to the next largest curve, which is also appropriate for the $(2, 2)$ FISP, and so on. As we might expect, the decision numbers decrease with M, since we need to select more points. Figure 19 shows the associated probabilities of calling all M maxima, which obviously decrease with M, but perhaps not by as much as one might expect. The decision numbers needed for the $(5, 3)$ FISP are shown in figure 20. We find that the probability of calling all three of the largest data points is 0.32, 0.60 and 0.79 for $r = 3, 4$ and 5 respectively, and $n = 100$.

Results for the (r, M) FISP with correlation and feedback

Since we now have a large parameter space to explore, we will restrict our attention to a case that seems relevant to the triads problem, namely $\alpha = 0.2$ (correlation over about five data points), $f = 0.5$ and $M = 3$, for various values of r. Figure 21 shows the three sets of decision numbers for $r = 10$, 15 and 20. We can see that, as r gets larger, the sets become closer to each other. This is in line with British Energy's current practice, where a single threshhold (decision number) is used to decide whether or not to call. Although one of the $r = 20$ sets is a little different from the other two, this is probably due to numerical error, since it contains the decision numbers for the $(18, 1)$ FISP, for which the probability of calling the maximum is greater than 0.99. Figure 22 shows the probability of calling the three largest data points as a function of r. This is greater than 0.9 when $r \geq 18$. Also shown for comparison are the same probabilities when there is no feedback and/or no correlation. Just as we found when $M = 1$, the effect of feedback on these probabilities is significantly greater than the effect of correlation. This is an aspect of the problem that could be investigated further.

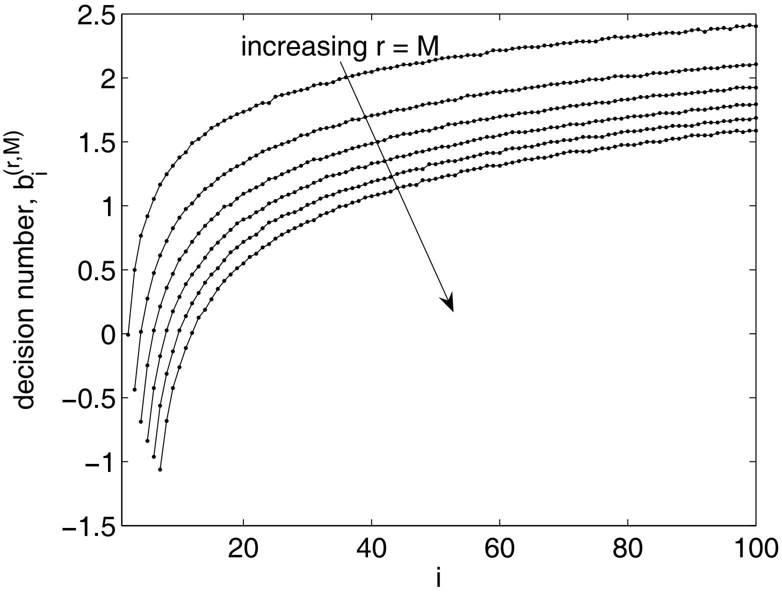

Figure 18: The decision numbers for the (M, M) Full-Information Secretary Problem with no correlation or feedback ($\alpha = 1,\ f = 0$), calculated by Monte Carlo simulation.

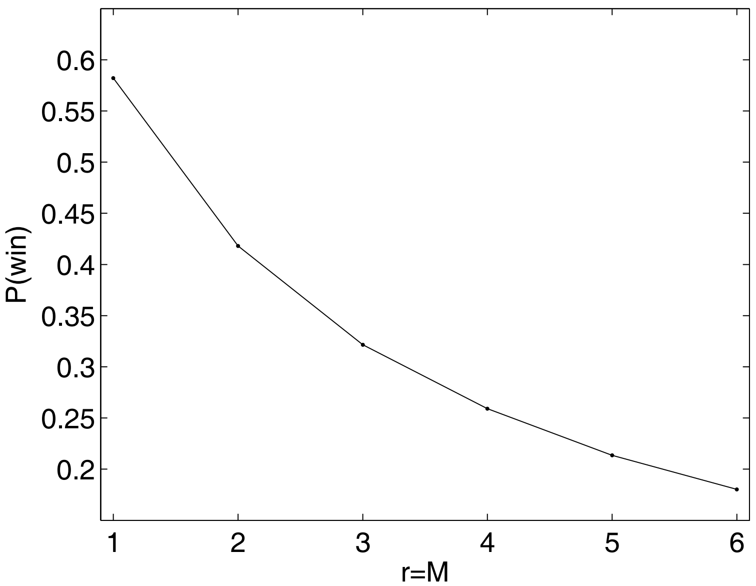

Figure 19: The probability of calling the M largest values for the (M, M) Full-Information Secretary Problem with no correlation or feedback ($\alpha = 1,\ f = 0$), calculated by Monte Carlo simulation for $n = 100$.

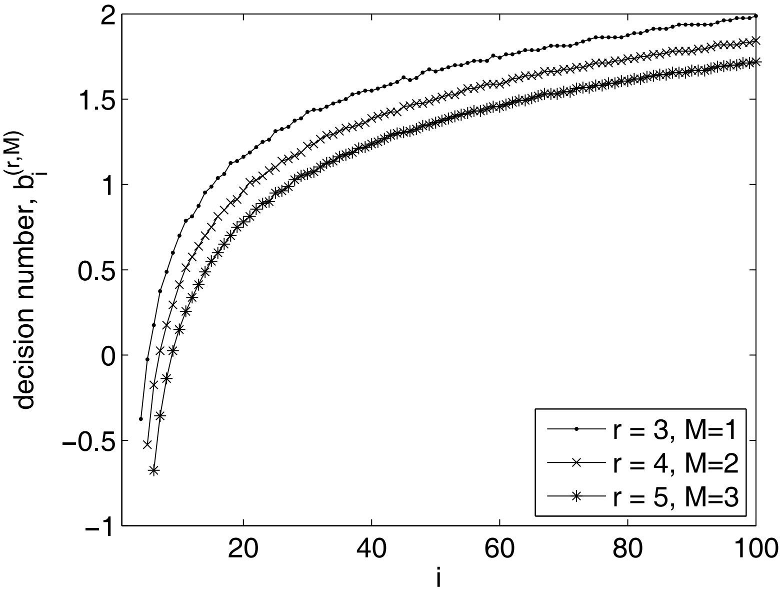

Figure 20: The decision numbers needed to solve the $(5, 3)$ Full-Information Secretary Problem with no correlation or feedback ($\alpha = 1,\ f = 0$), calculated by Monte Carlo simulation.

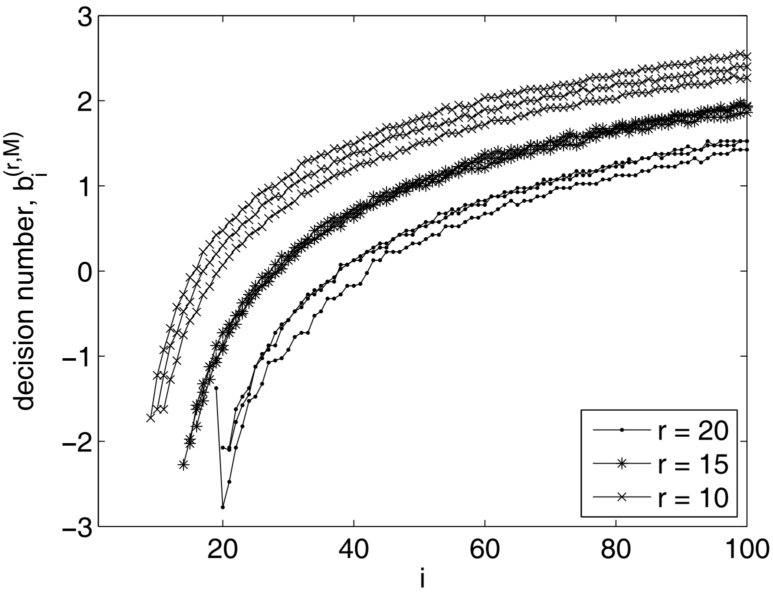

Figure 21: The decision numbers needed to solve the $(10, 3)$, $(15, 3)$ and $(20, 3)$ Full-Information Secretary Problems with correlation $\alpha = 0.2$ and feedback $f = 0.5$, calculated by Monte Carlo simulation.

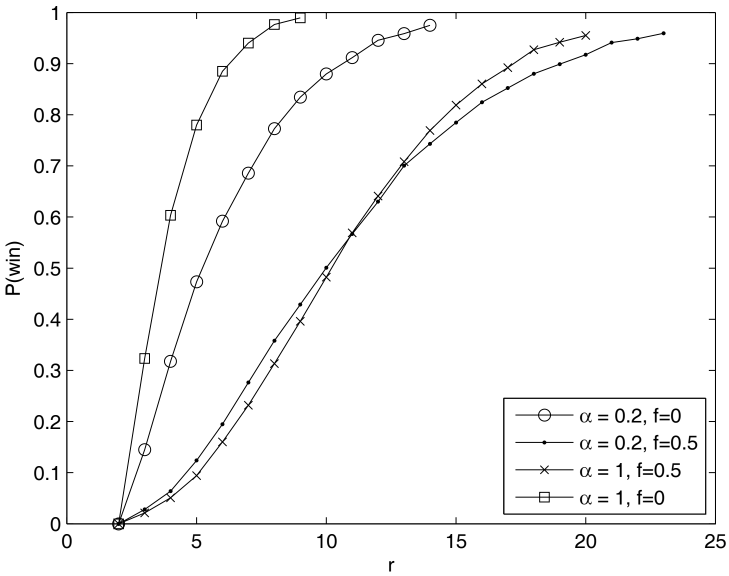

Figure 22: The probability of calling the 3 largest values for the (r, M) Full-Information Secretary Problem with correlation $\alpha = 0.2$ and feedback $f = 0.5$, calculated by Monte Carlo simulation for $n = 100$. In order to compare the effects of correlation and feedback, the results for $\alpha = 1$ and/or $f = 0$ are also shown.

5 Criteria based on analysis of historical data

As previously described, the current algorithm used by British Energy, implemented by TriFoS, is extremely effective. However, in this section we consider alternative or additional criteria that might be used for deciding whether to issue a Triad warning. Such criteria would involve rewriting the TriFoS algorithm and are therefore unlikely to be implemented; however, they are interesting in their own right.

Notation:

(1) We define the actual Peak Power consumption on a given day in year n as $C_a(i,n)$ with $1 \le i \le 120$ representing the date and n the year. Here we take n in the range $1 \le n \le 5$ corresponding to years 2001/02 to 2005/06 and dates from $1/11/n$ to $28/2/(n+1)$. Forecast Peak Power consumption $C_f(i,n)$ is defined similarly.

(2) From $C_a(i,n)$ and $C_f(i,n)$ we define $\tilde{C}_a(i,n)$ and $\tilde{C}_f(i,n)$ which are the ordered Peak Power consumptions, with $\tilde{C}_a(1,n) > \tilde{C}_a(2,n) > \cdots > \tilde{C}_a(120,n)$.

Notes:

(1) For technical reasons, historical figures are not available on every single day. The number of total Actuals and Forecasts are given in Table 3.

Year	Number of Actuals	Number of Forecasts
2001/02	60	58
2002/03	55	52
2003/04	60	57
2004/05	64	63
2005/06	64	64

Table 3: Number of Actual and Forecast Peak Power levels recorded.

(2) In 2005/06 Scotland was included for the first time. In order to compare $\tilde{C}_a(i,5)$ with $\tilde{C}_a(i,n)$, $n < 5$, we subtract an estimated consumption for Scotland using the estimate $C_a(i,5) - \frac{1}{4}\sum_{n=1}^{4} C_a(i,n)$.

(3) It is important when comparing consumption from year to year to attempt to make sure that we are looking at data collected in the same way. Hence we have only used $\tilde{C}_a(i,n)$ and $\tilde{C}_f(i,n)$ for i in the range $1 \le i \le 50$.

(4) We assume that TriFoS is able, or can be adapted, to produce a good forecast of the actual consumption $C_a(i,n)$ on a given day i, either with or without taking into account the effects of negative feedback. We also presume that this forecast is better than that given by National Grid. Although actual consumption and the National Grid forecast are clearly well correlated (with a correlation coefficient of 0.95; see Figure 23), we find that the National Grid forecast minus the actual consumption, for the year 05/06, has a mean of 881 with a standard deviation of 556. Least-squares fit of a straight line gives a slight slope, but is not very different to the line with height equal to the mean (see again the figure). The higher National Grid estimate is partially but not entirely due to negative feedback.

It is clear from this graph that a year-by-year comparison on a given day will have quite a large random element in it.

Aims:

(1) We want to attempt to remove as much as possible of the random element in the figures, and to do that we compare the ordered forecasts, $\tilde{C}_a(i,n)$.

(2) We then want to see if we can make a good estimate of when we should issue a Triad warning, assuming that a good estimate of the actual consumption is available.

(3) The days close to an actual Triad are excluded because of the ten-day rule. If we filter these out then the top three in the amended list are the Triads. The i associated with the $\tilde{C}_a(i,n)$ of the Triads form a subsequence relating where the Triads appear in the ordered list. We note here that for the available data all Triad days have $i < 20$, as is given in Table 4.

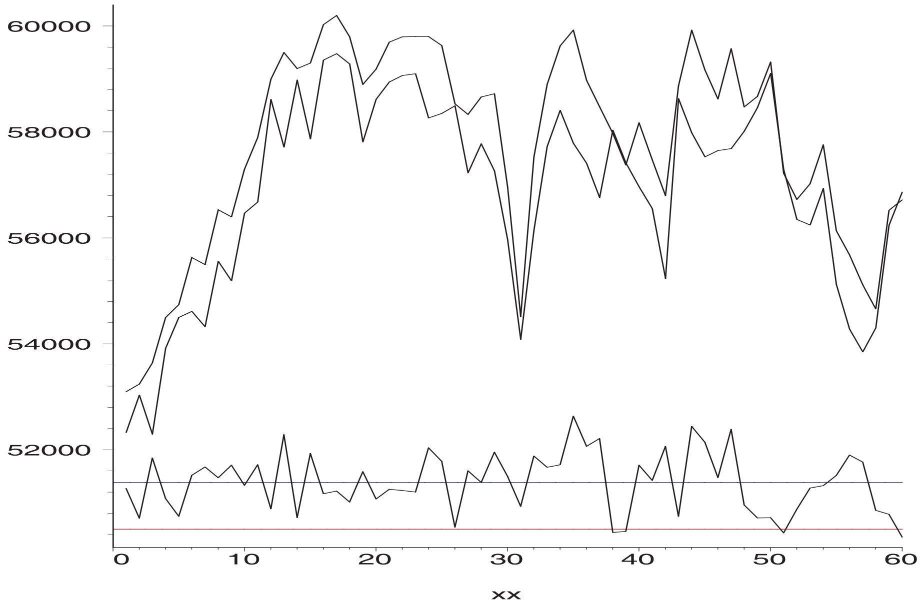

Figure 23: National Grid forecast and actual consumption for 2005/06. The top two curves are respectively the forecast and the actuals. The lower curve is the difference, with the lower line representing the axis and the upper line the mean of 881.

Year	Triad 1	Triad 2	Triad 3
2001/02	1	2	17
2002/03	1	2	14
2003/04	1	4	11
2004/05	1	4	14
2005/06	1	9	12

Table 4: The ordinal i associated with each Triad.

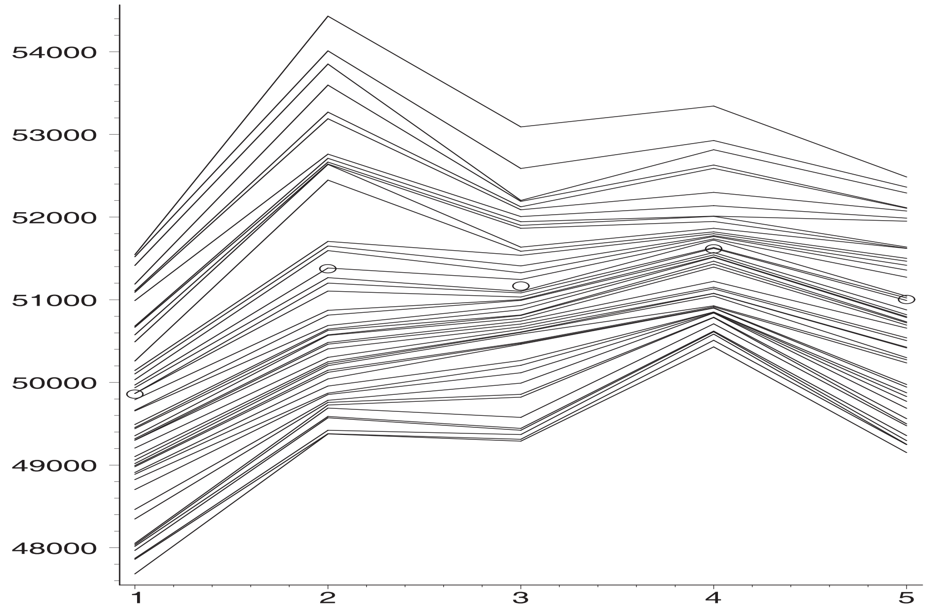

Figure 24: Graph joining the points $\tilde{C}_a(i,n)$ for fixed i and $1 \leq n \leq 5$, with the Scottish average subtracted in the year 2005/06.

Figure 24 is a piecewise continuous graph joining the points $\tilde{C}_a(i,n)$ for fixed i with $1 \leq n \leq 5$, where the estimated Scottish average of 6991 is subtracted to make data uniform from year to year. The top 40 of these graphs are shown and the circles represent the average of the top 40 for a given year. Suppose we define the signature of each graph, for example $(+,-,+,-)$, as the sign of the piecewise derivative, that is the sign of the slope of the graph between year n and $n+1$ for a fixed value of i. It is clear that $\tilde{C}_a(1,n)$ dominates, in that its signature is shared by most of the other graphs as we vary i. The average also has this same signature and thus is a good indicator of the shape of each graph for fixed i.

From these observations, the average is not the only feature that determines the top structure in the ordered list: the signature is also an important feature. In order to predict this structure accurately it is therefore good to use at least two parameters. For issuing Triad warnings, the objective is to predict the level of consumption $\tilde{C}_a(20,n)$ from the previous years' data and any data so far available for the current year to date. The average of $\tilde{C}_a(20,n)$ still shows a sizable random element from year to year as is seen from the circles in Figure 24 and in Table 5. While the difference shows that the average is well approximated by the average of the top 40, as we would expect, there is a larger random yearly element.

We now turn our attention to the best parameters to determine the values of $\tilde{C}_a(i,n)$ in the ordered list. As we indicated earlier we require two parameters and this suggests that we should be looking at a straight line fit.

Figure 25 plots $\tilde{C}_a(i,n)$, for n fixed and i in the range $1 \leq i \leq 40$, and the best straight line fit. All 5 years are superimposed: it is not really important to know which year is which, it is the distribution of errors that is important. Figure 26 shows the error of the

Year	$\tilde{C}_a(20,n)$	Average	Difference
2001/02	49791	49859	-68
2002/03	50872	51376	-504
2003/04	51001	51168	-167
2004/05	51572	51616	-44
2005/06	57868	57995	-127

Table 5: Comparison between $\tilde{C}_a(20,n)$ and the average of the top 40.

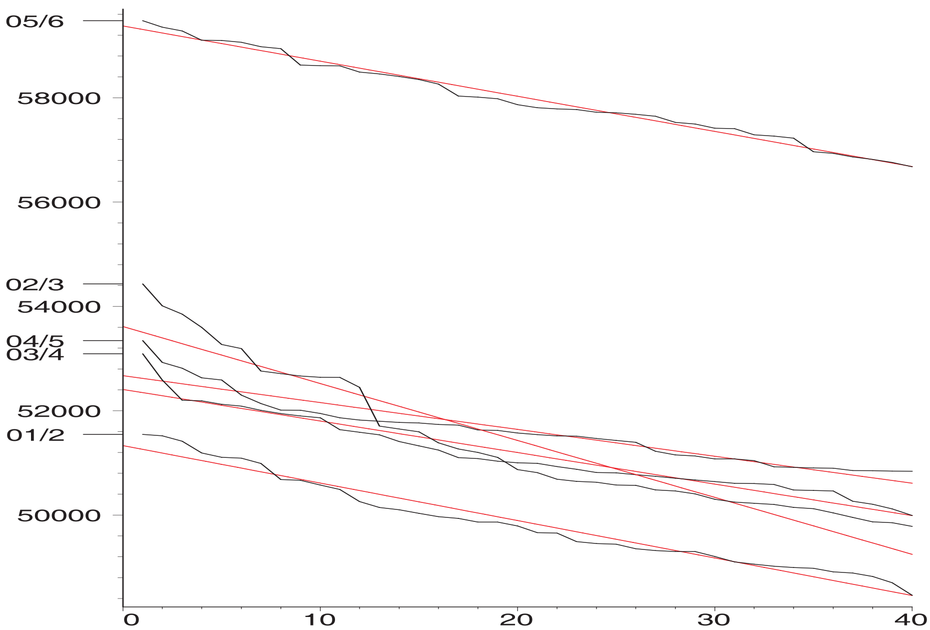

Figure 25: Graphs joining the points $\tilde{C}_a(i,n)$ for a fixed n, with the best straight line fit.

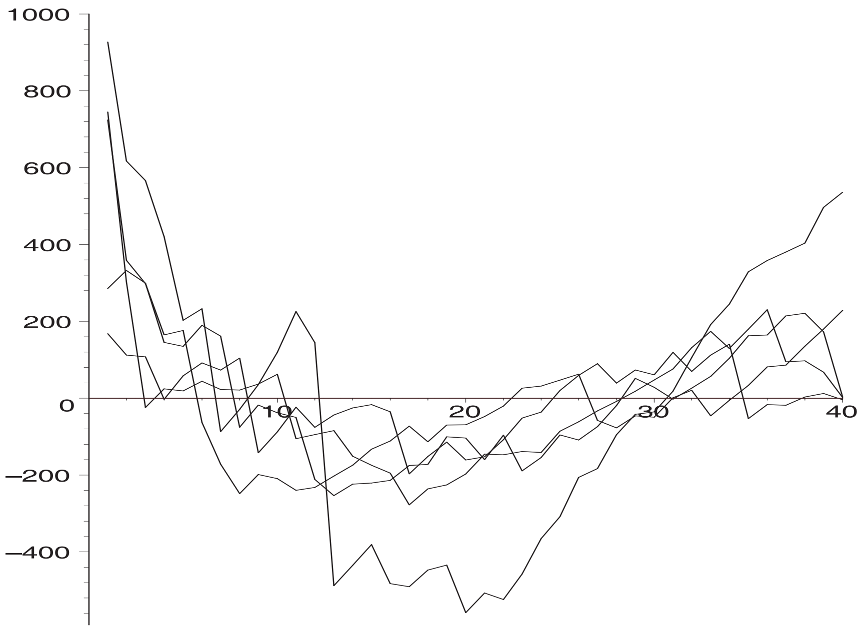

Figure 26: Errors in the straight-line fit of the top 40.

Year	$\tilde{C}_a(15,n)$	Linear fit	$\tilde{C}_a(20,n)$	Linear fit	$\tilde{C}_a(25,n)$	Linear fit
2001/02	50033	50059	49791	49742	49440	49424
2002/03	51595	51493	50872	50980	50576	50466
2003/04	51326	51261	51001	51020	50810	50778
2004/05	51766	51761	51572	51593	51431	51425
2005/06	58349	58244	57868	57938	57712	57631

Table 6: Comparison between $\tilde{C}_a(i,n)$ and the linear fit for $i = 15$, 20 and 25.

linear fit, equal to the difference between $\tilde{C}_a(i,n)$ and the linear fit. All five years show that the trend errors are positive at the start and finish, that is the highest and lowest consumption is underestimated by this fit, and negative in the middle ($15 \leq i \leq 25$). It should be noticed that the range of errors in Figure 26 is approximately $(-450, 1000)$ while it is reduced to approximately $(-450, 100)$ for $15 \leq i \leq 25$. However the range of each individual year is much less. Figure 27 and Table 6 emphasise this by comparing the consumption with a linear fit of the consumption in the range $15 \leq i \leq 25$. We see that this fit predicts the consumption in the range to within an error of ± 110.

Summary and analysis of data

The Peak Power consumption on a given day and year, $C_a(i,n)$, varies from year to year and exhibits considerable random fluctuation. The ordered Peak Power consumption,

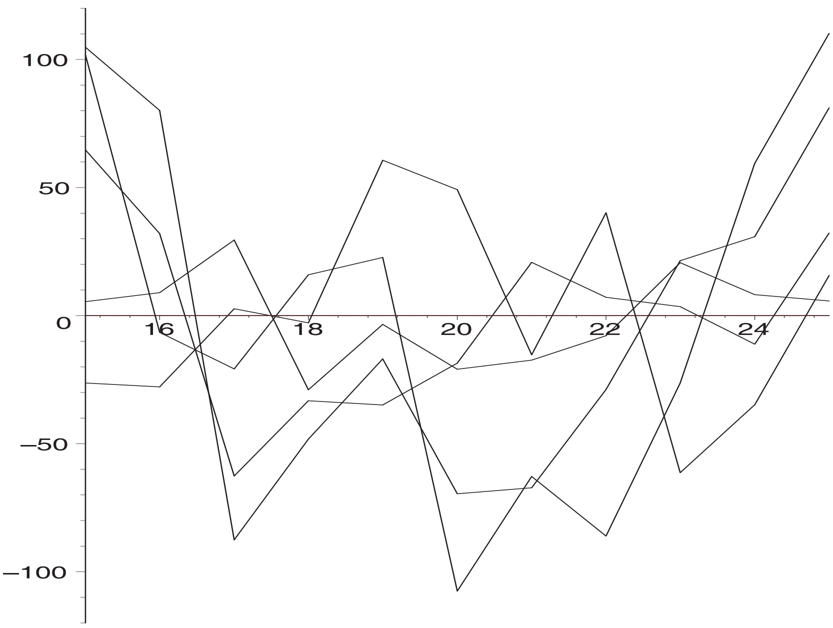

Figure 27: Errors in a straight-line fit restricted to $15 \leq i \leq 25$.

$\tilde{C}_a(i,n)$, shown in Figure 24 displays much less fluctuation from year to year. We also show in Figure 25 that for a fixed year the ordered Peak Power consumption can be well approximated by a straight line. However we should remember that Figure 25 also shows that there is considerable variation of this fit from year to year. The error in the straight line fit of the top forty is shown in Figure 26, while the goodness of fit of the top 15 to 25 is shown in Figure 27, which shows that the absolute error is less than 110. We should, of course, point out that while it is easy to fit this straight line "after" the event, that is at the end of the year, it is more difficult to predict this fit at the beginning of the year. However tables 4 and 5 show that if we are able to predict the value of $\tilde{C}_a(20,n)$ we should be able to predict, and thereby issue a warning for, all the likely Triad events. If we do issue warnings for all of the top 20 then the problem of negative feedback should almost disappear, because we expect to have caught all the Triads.

Predicting the value of $\tilde{C}_a(20,n)$

While predicting the value of $\tilde{C}_a(20,n)$ is the main aim, we suggest that the best strategy is to predict the linear fit. This requires two parameters, either the mean and the slope, or perhaps better using two points. We suggest that we take $\tilde{C}_a(15,n)$ and $\tilde{C}_a(25,n)$. However we should remember that we are not actually trying to predict $\tilde{C}_a(15,n)$ and $\tilde{C}_a(25,n)$ but merely the straight line, which will then give not only give an estimate of $\tilde{C}_a(20,n)$ but also an estimate of the error in the predicted value. We also note that a quadratic fit requires three points and a cubic four points which would be very difficult to forecast accurately.

We can get some prior estimate from Table 6, taking into account that Scotland is included in the consumption figures now. What is then required is to update these estimates in the light of the consumption figures of early November, before the likely period of the first Triad date, and throughout the year as more data comes in. We have not attempted to do this. However we should remember that an estimated linear fit will not have as good error bounds as the actual fit and this should be taken into account when deciding whether to issue a warning. The best suggestion is for this alternative method to be used in a trial period running simultaneously with the present TriFoS method and comparing the results at the end of the year to see which gives the better results.

6 Conclusions and further work

We began the study by conducting an analysis of the errors in National Grid's consumption forecasts, which revealed a statistically significant dependence on whether or not Triads calls had been issued by British Energy. This result confirms that negative feedback is present in practice. We then went on to consider two broad ways of approaching the question of identifying Triads as they occur, based on extensions of the Full-Information Secretary Problem, and a direct analysis of historical data.

Methods based on the Full-Information Secretary Problem

We have shown that by studying $(r, 3)$ Full-Information Secretary Problem with feedback and correlation we can gain some insight into the Triads problem, and that our results, which suggest that about 18 calls are needed to be able to find the three largest data points 90% of the time, are in line with the experience of British Energy. Our results also indicate that negative feedback may be a more important issue than the correlated nature of the data. We were also able to derive some new analytical results for the $(r, 1)$ FISP and $(1, 1)$ FISP with feedback.

In order to make these simulations more realistic, we would need to include the effect of the ten-day window, and add some uncertainty into the prediction. For example, this uncertainty could take the form of a value chosen from a normal distribution and added to each data point *after* the decision to call or not has been taken. The optimal, or at least a good, strategy then needs to be more complicated. We would wish to call data points whose predicted value is not one of the three largest so far, but is sufficiently large that uncertainty may make them so.

In addition, we note that the Triad problem has two more subtle aspects in practice. Firstly, the aim should be to maximise the expected number of triads called, not to maximise the probability of calling all three triads. However, when the probability of the latter is close to unity, the expected number of triads called is sure to be close to three. Secondly, British Energy would like to minimise the number of calls made. This has not been addressed here, as we have assumed that a fixed number of calls, r, is available. It would be worth looking at the results again to assess the expected number of calls, which is likely to be less than r when r is large, but we have not attempted this here.

Methods based on analysis of historical data

If peak daily consumptions through the winter are placed in decreasing order, then the Triad days always seem to fall within the top 20. Since British Energy has at least 20 calls available, calling the top 20 will catch the actual Triads (as above, this line of thinking does not attempt to minimise the number of calls made). We therefore looked at ways of predicting the twentieth highest daily consumption and found that, within each year, a simple linear fit to the ordered daily peak consumptions may be sufficient. Although there is significant variation from year to year, it may be possible to use actual data from early November in each year to parameterise the linear fit for that year before any of the Triads occur.

A Optimal strategy for the $(1, 1)$ FISP

Following Gilbert and Mosteller [2] for the case of data points drawn from $U(0, 1)$, consider the position when the current data point, x, is the largest so far, and we have yet to call a maximum, with k data points still to come. If we call a maximum, the probability that we are correct is equal to the probability that the remaining k data points are less than x, which is x^k. If we don't call a maximum, but choose the next value higher than x, the probability that we win when there are j data points remaining that are higher than x is equal to the probability that the highest of these j data points comes first in the sequence,

$$\frac{1}{j}\begin{pmatrix} k \\ j \end{pmatrix} x^{k-j}\,(1-x)^j\,,$$

where

$$\begin{pmatrix} k \\ j \end{pmatrix} \equiv \frac{k!}{j!(k-j)!}.$$

The appropriate decision number for the current data point is the value of x at which the probability of winning by calling a maximum and the probability of winning by not calling a maximum are equal. This gives us

$$x^k = \sum_{j=1}^{k} \frac{1}{j}\begin{pmatrix} k \\ j \end{pmatrix} x^{k-j}\,(1-x)^j\,,$$

or, defining $z = (1-x)/x$,

$$1 = \sum_{j=1}^{k} \frac{1}{j}\begin{pmatrix} k \\ j \end{pmatrix} z^j. \tag{7}$$

This equation is straightforward to solve numerically, and gives $x = 1/(1+z) = b_{k+1}$. For example, when $k = 1$, (2) reduces to $z = 1$, and hence $b_2 = 0.5$, as we would expect when there is just one data point remaining. We also note that the right hand side can be written as an integral, so that (2) becomes

$$1 = \int_0^z \frac{(1+s)^k - 1}{s}\, ds. \tag{8}$$

Although this formulation has no advantage over (2) for numerical evaluation, we can see that, when $k \gg 1$, $z = O(1/k)$, so that, if we define $z = \bar{z}/k$, $s = \bar{s}/k$, (8) becomes, at leading order as $k \to \infty$,

$$1 = \int_0^{\bar{z}} \frac{e^{\bar{s}} - 1}{\bar{s}} d\bar{s}, \tag{9}$$

which has the unique solution $\bar{z} = \bar{z}_0 \approx 0.80435$. This shows that

$$b_{i+1} \sim 1 - \frac{\bar{z}_0}{i} \quad \text{as } i \to \infty. \tag{10}$$

B An almost-optimal strategy for the $r, 1$ FISP

In the notation of section 3.4 Using these ideas, we can write

$$P_2 = \sum_{j=R}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) x^{k-j} (1-x)^j \left(1 - \sum_{I=R}^{k} \sum_{J=R-1}^{I-1} P_{jIJR} \right),$$

where P_{jIJR} is the probability that the largest of the remaining j data points larger than x is the Ith, with J calls required for the preceding $I-1$ data points.

Since, of the $j!$ permutations of the remaining j data points larger than x, there are $(j-1)!$ with the maximum being the Ith, we have

$$P_{jIJR} = \frac{1}{j!}(j-1)! \frac{1}{(I-1)!} \left[\begin{array}{c} I-1 \\ J \end{array} \right] = \frac{1}{j} \frac{1}{(I-1)!} \left[\begin{array}{c} I-1 \\ J \end{array} \right],$$

and hence the probability of winning by choosing the ith data point is

$$\begin{aligned} P_1 + P_2 \;\; = \;\; & \sum_{j=0}^{R-1} \left(\begin{array}{c} k \\ j \end{array} \right) x^{k-j} (1-x)^j \\ & + \sum_{j=R}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) x^{k-j} (1-x)^j \left(1 - \frac{1}{j} \sum_{I=R}^{k} \sum_{J=R-1}^{I-1} \frac{1}{(I-1)!} \left[\begin{array}{c} I-1 \\ J \end{array} \right] \right). \end{aligned}$$

Similar arguments can be used to determine the probability of winning by continuing, which is

$$\sum_{j=1}^{R} \left(\begin{array}{c} k \\ j \end{array} \right) x^{k-j} (1-x)^j + \sum_{j=R+1}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) x^{k-j} (1-x)^j \left(1 - \frac{1}{j} \sum_{I=R+1}^{k} \sum_{J=R}^{I-1} \frac{1}{(I-1)!} \left[\begin{array}{c} I-1 \\ J \end{array} \right] \right).$$

After equating these probabilities, and noting that there is considerable cancellation between the two, we find that the decision number is the solution of the equation

$$1 = \sum_{j=R}^{k} \frac{1}{j} \left(\begin{array}{c} k \\ j \end{array} \right) z^j \sum_{I=R}^{j} \frac{1}{(I-1)!} \left[\begin{array}{c} I-1 \\ R-1 \end{array} \right], \tag{11}$$

where, as usual, $z = (1-x)/x$. This reduces to (2) when $R = 1$.

References

[1] Benjamin, A.T. and Quinn, J.J., 2003, *Proofs that Really Count: The Art of Combinatorial Proof*, Dolciani Mathematical Expositions.

[2] Gilbert, J.P. and Mosteller, F., 1966, Recognizing the maximum of a sequence, *J. Am. Stat. Ass.*, **61**, 35-73.

[3] Hlynka, M. and Sheahan, J.N., 1988, The secretary problem for a random walk, *Stoch. Proc. App.*, **28**, 317-325.